

Analysis of Current Metadata Descriptions of von Asch Items.

The items Georg Thomas von Asch (1729 – 1807) sent to the Göttingen University between 1771 and 1806 included artifacts from nature as well as ethnographic objects, books, maps, coins, medals etc. Most of these items were incorporated into the Royal Academic Museum, but when the museum was closed after the death of Johann Friedrich Blumenbach (1752 – 1840), the collections of the museum were distributed among the departments of the Göttingen University and so were the items of the provenance von Asch: The skulls, Baron von Asch had sent to Blumenbach went to the Department of Anatomy, the coins and medals to the Coin Collection of the Archaeological Institute, the ethnographic objects to the Ethnographic Collection, geological items to the different geological collections (now in the Collections of the Geoscience Center), print holdings to the Art Collection, zoological objects to the Zoological Museum and seeds and plants to the Botanical Collections. This situation should continue for another two hundred years. At the end of the 20th / beginning of the 21st century, the University Collections started to catalogue some of their artifacts in databases, most of them used in-house solutions for software and the metadata schemas. As a result, the data of the different Collections is rarely interoperable with each other and may be merged only on a generic level.

In the context of the ASCH project we want to develop a model that may merge data describing these different sorts of artifacts. Therefore we examined the data of those Collections owning items of the von Asch provenance. Because the descriptions of von Asch items are in most cases part of a bigger data set, the analysis applied to the whole data set using i. a. Open Refine, which is especially useful for the validation and enrichment of the CSV files most of the Collections provided. We did an analysis of the following datasets:

- The Ethnographic Collection: The Collection uses an Adlib database for the description of artifacts. As a software developed especially for the museum community, Adlib follows standards like Spectrum and LIDO.¹ However, for the data analysis we received the data in a proprietary XML Schema.
- Collections of the Geoscience Center: For most of the items the Geoscience Center uses in-house solutions for databases and metadata schemas. But items of the scientific collection of Johann Friedrich Blumenbach, which include items sent by Baron von Asch, are recorded in a cross-domain database developed by the Blumenbach-Online project².
- Collections of the Institute of Archaeology: Most of the databases are in-house solutions. One of these solutions is an Open Office database describing the medals von Asch sent to the University of Göttingen which we analyzed concerning the reuse of the data in the context of the ASCH project. In addition to this collection we looked at the KENOM database, an easyDB database provided by the VZG³ for the recording of coins and other currencies

¹ <http://www.adlibsoft.de/produkte/museum-collection-management-software>

² <http://www.blumenbach-online.de/>

³ <https://www.gbv.de/Verbundzentrale/serviceangebote/kenom>

across the borders of collections and organizations. This database provides data as a cross-domain LIDO export⁴ and is used for the recording of the Coin Collection of the Institute of Archaeology. It may also be used in the future for the structured description of the coins which von Asch sent to the University.

- Museum of Zoology: The items of the Museum of Zoology are described in an in-house database (based on filemaker). The dataset we analyzed was a CSV-file of about 23.000 records.
- The Blumenbach Skull Collection: This collection includes the skulls von Asch sent to Göttingen which are recorded in the cross-domain database developed by the Blumenbach-Online project.
- Arts Collection: The metadata that will be used in the ASCH project was recorded in kuniweb, an easyDB database also provided by the VZG and used for the cataloguing across domains and organizations of items especially from museums and university collections. The LIDO export schema for this database is still work in progress. Because we could not get a LIDO dataset, we examined the cataloguing interface.
- Göttingen State and University Library: Items of the von Asch provenance are described in the Union Catalogue GBV. Most of these items are digitized and the metadata is available in the DFG-Viewer compliant METS/MODS schema. These digitized items are also made available in the Deutsche Digitale Bibliothek [e.g. <https://www.deutsche-digitale-bibliothek.de/item/PIG3VX7YSC5G3YSU43VAETISTGVEBU6>], but are not described as items of the von Asch provenance there.

The following table compares the data provided by the collections with the requirements we defined during the Use Case Work Package of the ASCH project. We used the following marks:

✗ not compliant

(✓) partly compliant

✓ compliant

⁴ <https://www.gbv.de/Verbundzentrale>

Resource	Superclass of all classes used in the model	Entity				
[Requirement 6]	Resources must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifier	Unique and persistent identifiers are only used in those data sets that use cross-collection databases. Therefore this requirement can be fulfilled only partially.		
[Requirement 7]	Resources must be identified by a name/appellation that allows to distinguish between resources of the same type	Property	Title	In the most cases, the described resources can be distinguished from other resources by using a distinct appellation.		
[Requirement 8]	The nature of a resource must be described using 1-n controlled values.	Property	Type	Generally, the type of the resource is described, mostly by using values of controlled vocabularies (classifications or taxonomies).		
[Requirement 9]	Resources can be interlinked with 0-n other resources.	Occurrence	Repeatability	Generally, an unlimited repeatability of fields is impossible.		
[Requirement 10]	Resources must be interlinked with each other using unique, machine readable and persistent identifier.	Values	Identifier	Unique, persistent and machine readable identifiers occur in cross-collection databases only.		
[Requirement 11]	The nature of the relation between two resources can be specified using controlled values.	Property	Relations			
[Requirement 12]	The relation between two resources can be specified by the time this relation was valid.	Property	Relations			

[Requirement 71]	Topics of information resources must be described using controlled values	Property	Subject Headings	This requirement is relevant for the Art History only because here the possibility of using controlled values is given by the classification system Iconclass.			✓
[Requirement 72]	Resources used as access points must be identified as instances of one or more classes.	Property	Type of Resource	This requirement is fulfilled in cross-collection databases. The Zoology is using the ITIS, an internationally common vocabulary.			✓
Metadata Set	The machine readable description of a single resource by statements.	Entity					
[Requirement 1]	Resource descriptions must be machine readable.		Metadata	The bigger part of the data is structured and machine readable.			✓
[Requirement 2]	Resource descriptions must be compliant to the one-to-one principle		Metadata	The distinction of resource descriptions is unambiguous because one single resource is described by one single record.			✓
[Requirement 3]	A resource description must provide all attributes necessary to identify the resource and distinguish it from other resources of the same type.	Property	Description	Given in the checked data.			✓
[Requirement 4]	Resource descriptions must be identified by identifiers that are unique, machine readable and persistent.	Property	Administrative Metadata	Beyond the respective systems, the used record identifiers are not unique.			✗
[Requirement 5a]	Creation of resource descriptions must be documented by time stamps.	Property	Administrative Metadata	Only partially the case in the checked data.			(✓)
[Requirement 5b]	Modification of resource descriptions must be documented by time stamps.	Property	Administrative Metadata	Only partially the case in the checked data.			(✓)
[Requirement 63]	Resource descriptions must include information about the reusability of these descriptions (e.g. license	Property	Administrative				✗

	information)		Metadata			
[Requirement 64]	Resource descriptions must be compliant to a cross-domain or domain-specific metadata standard.			Until now, most of the databases do not provide standard conform data exports. Only the data documented in KENOM is available as LIDO data and therefore it is corresponding to a domain specific standard. Currently, a LIDO export for kuniweb is in progress.	(✓)	
[Requirement 65]	Resource descriptions must refer to the used metadata standard(s)	Property	Administrative Metadata	In the LIDO data reference is made to the LIDO schema.	(✓)	
Item	A real world thing in a collection	Entity				
[Requirement 13]	Item descriptions can be interlinked with evidence.	Property	Relations	An interlinking with digital evidences is to be found in the Arts Collection only, and here only for the provenance von Asch.	X	
[Requirement 20]	Item descriptions must be interlinked with 1-n events in the lifecycle of the item.	Property	Events	This requirement is fulfilled concerning the data available as LIDO export. Generally, this is not the case with the other datasets.	(✓)	
[Requirement 75]	Items must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifiers	Generally, items are identified by inventory numbers that are unique within a certain collection. Only the data in KENOM and kuniweb provide unique and persistent identifiers usable in machine processing by using handles.	(✓)	
[Requirement 76]	Items must be identified by a name or appellation.	Property	Title	Generally, the items have an unambiguous appellation.	✓	
[Requirement 42]	For items a digital representation/copy can be available.	Property	Digital Representations	Digital representations of the items are recorded especially in the cross-collection databases. This is rarely the case concerning the other datasets.	(✓)	
[Requirement 44]	Items can be represented by digital representations in different resolutions.	Property	Digital Representations	In the cross-collection databases, it is possible to provide digital representations in different resolutions.	(✓)	
[Requirement 51]	Item descriptions must be assigned to 1-n collections.	Property	Relations	In the cross-collection databases, it is possible to assign an item to more than one collection. This is not intended for most of the other databases.	(✓)	

[Requirement 52]	All items of a collection must be interlinked with the collection using unique, machine readable and persistent identifiers.	Property	Collections	The interlinking using unique and persistent identifiers is possible in the cross-collection databases only.	(✓)
[Requirement 77]	Item descriptions must interlink with information about the current location of the item.	Property	Location	In the cross-collection databases, unambiguous location information is recorded. The location information recorded within the other collections are usable only internally.	(✓)
[Requirement 59]	Item descriptions must provide machine readable information about the accessibility and usability of items.	Property	Rights	Generally, not existing.	X
Evidence	A resource proving the reliability of a statement about a resource.	Entity			
[Requirement 15]	Evidences can be machine readable.		Metadata	Until now, the evidences are digitized marginally and are generally not available as structured full texts.	X
[Requirement 16]	The nature of an evidence must be described by 1-n controlled values	Property	Type of the Resource	Until now, the evidences are rarely be indexed.	X
[Requirement 17]	Information about resources in an evidence must be identified by unique identifiers.	Property	Identifier	Until now, the evidences are digitized marginally and are generally not available as structured full texts..	X
[Requirement 18]	Information about the nature of a resource in an evidence must be identified by 1-n controlled values.	Property	Type of the Resource	Until now, the evidences are digitized marginally and are generally not available as structured full texts..	X
[Requirement 78]	Evidence descriptions must provide information about the accessibility and reusability of the resource.	Property	Rights	Generally, this is not the case.	X
Event	An activity in the lifecycle of a	Entity			

	resource						
[Requirement 79]	Event descriptions can be interlinked with evidence about the event (e.g. documents)	Property	Relation	Until now, relations between events and evidences are not represented.			X
[Requirement 21]	Event descriptions must be machine readable.		Metadata	The event descriptions within the LIDO data are machine readable.			(✓)
[Requirement 22]	Event descriptions must show the nature of the relation between items, agents, places and/or time.	Property	Relations	This requirement is fulfilled concerning the data that can be provided as LIDO export.			(✓)
[Requirement 23]	An event in the lifecycle must be related to 1-n items.	Property	Relations	This is possible concerning the LIDO exports, but has not been implemented, so far.			X
[Requirement 24]	An event in the lifecycle of an item must be related to 0-n agents	Property	Relations	This requirement is fulfilled concerning the data that can be provided as LIDO export.			(✓)
[Requirement 25]	The function/role of an agent during an event must be specified using controlled values (e.g. creator, dealer, collector).	Property	Relations	This has been implemented within the LIDO data, partially.			(✓)
[Requirement 26]	An event in the lifecycle of an item must be related to 0-n places.	Property	Relations	This has been implemented within the LIDO data, partially.			(✓)
[Requirement 27]	An event in the lifecycle of an item must be related to 0-n date information.	Property	Relations	This requirement is fulfilled concerning the data that can be provided as LIDO export.			(✓)
Time	A time span related to a resource via an activity or as a topic	Entity					
[Requirement 29]	Date and time indication must be machine readable.	Datatype	Values	Given within the cross-collection databases.			(✓)

[Requirement 30]	Date information must be compliant to date type standards.	Datatype	Values	Given within the cross-collection databases.			(✓)
[Requirement 80]	A human readable form of date information can only be used in addition to a machine readable form.	Value	Values	Possible within the cross-collection databases			(✓)
[Requirement 81]	A human readable form of date information can use values of a controlled vocabulary.	value	Values	Until now, not intended.			X
Agent	A person, organization or group related to a resource via an activity or as a topic.	Entity					
[Requirement 32]	Agents must be identified by name and biographical information.	Property	Description	In the cross-collection databases, this is provided by linking to authority data.			(✓)
[Requirement 33]	Agents must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifiers	In the cross-collection databases, this is provided by linking to authority data.			(✓)
[Requirement 34]	Agent descriptions can interlink with resources related to the agent.	Property	Relations	Until now, not intended.			X
[Requirement 35]	Agent descriptions must be machine readable.		Metadata	In the cross-collection databases, authority data are used that are available as machine readable data.			(✓)
[Requirement 36]	Agent descriptions can interlink with places related to the agent.	Property	Relations	Has been realized by the use of authority data within the cross-collection databases.			(✓)
[Requirement 37]	The relation between an agent and a place can be specified by the character of the relation (e.g. visitor or resident)	Property	Relations	Until now, not intended.			X

[Requirement 38]	The relation between an agent and a place can be specified by the time this relation was valid.	Property	Relations	Until now, not intended.			X
Place	A geographic location related to a resource via an activity or topic.	Entity					
[Requirement 39]	Places must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifier	Has been realized within the cross-collection databases.			(✓)
[Requirement 40]	Places must be identified by machine readable geo-information / geographic coordinates	Datatype	Value	Has been realized within the cross-collection databases.			(✓)
[Requirement 41]	Place descriptions can include 1-n place appellations.	Property	Title	Has been realized within the cross-collection databases.			(✓)
Digital Representation	A digital resource depicting an item	Entity					
[Requirement 43]	A digital representation/copy must enable the identification of the item it represents.		Digital Representations	Not all digital representations do fulfill this requirement.			(✓)
[Requirement 45]	The nature of a digital representation must be described using a controlled value.	Property	Type of the Resource	Has been realized within the cross-collection databases.			(✓)
[Requirement 60]	Descriptions of digital representations must provide information about the reusability of the digital representation (e.g. license).	Property	Rights	In the cross-collection databases, partially the photographer is named, but license information is not provided, until now.			(✓)
[Requirement 61]	Descriptions of a digital representation must provide contact	Property	Rights	Until now, not existent.			(✓)

	information related to the usability of a digital copy.					
[Requirement 82]	Digital representations must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifiers	In the cross-collection databases, realized by using handle.		(✓)
Collection	An aggregation of items:	Entity				
[Requirement 46]	Collection descriptions must be machine readable.		Metadata	When collection descriptions are available, they are machine readable.		✓
[Requirement 47]	Collections must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifiers	Generally, the collections of the Göttingen University have an ISIL identifier.		✓
[Requirement 48]	Collections must be identified by an explicit name.	Property	Title	The appellations of the collections are unambiguous.		✓
[Requirement 49]	Collection descriptions must include information about the location/place of the collection.	Property	Relations	The location information of the collections is provided via the ISIL in a machine readable form.		✓
[Requirement 50]	Collection descriptions must be compliant to a cross-domain metadata standard.		Standards	The data do not provide standardized descriptions of the collections.		X
Statement	A predication about an item.	Entity				
[Requirement 31]	A statement can be interlinked with date information.	Property	Date	Until now, an interlinking between statements about objects and statements about when the statements had been expressed is not possible.		X
[Requirement 53]	All statements must be repeatable.	Property	Repeatability	In the databases, an unlimited repeatability of existing fields is not allowed.		X

[Requirement 73]	Statements can be interlinked with information about their origin.	Property	Relations	Until now, an interlinking between statements about objects and statements about the origin of this information is not possible.		X
[Requirement 74]	Statements can be interlinked with resources.	Property	Relations	Until now, an interlinking between statements about objects and other resources is not possible.		X
[Requirement 54]	Statements can be interlinked with resources proving their reliability.	Property	Relations	Until now, an interlinking between statements about objects and evidences proving these statements is not possible.		X
[Requirement 55]	Inconsistency of statements is allowed if the metadata refers to the origin of the statements.	Validation	Repeatability	Until now, not possible.		X
[Requirement 86]	A statement can related to 0-n places.	Property	Relations	Until now, an interlinking between statements about objects and the related place, where the statement had been expressed, is not possible.		X
Holding	The place an item is located, subclass of place	Entity				
[Requirement 83]	Holdings must be identified by identifiers that are unique, machine readable and persistent	Property	Identifiers	Kuniweb enables an unambiguous and machine readable description of the location of the holding institution.		(✓)
[Requirement 84]	Holdings can be identified by machine readable geo-information / geographic coordinates.	Property	Value	In kuniweb, the location information is linked to geonames, so geographic coordinates are provided for the location, at least.		(✓)
[Requirement 85]	Holding descriptions must include 1-n place appellations.	Property	Value	Generally, the appellation of the location is not repeatable.		X
[Requirement 58]	Holding descriptions must provide contact information.	Property	Relations	Until now, ignored.		X
Concept	A term from a controlled vocabulary	Entity				
[Requirement 66]	Controlled values used for the description of resources must be	Value	Interoperability	In the cross-collection databases, cross-collection standards are used, e. g. GND or Iconclass. In the local developed databases as well as in the proprietary systems, mainly discipline-specific		✓

	values of a domain-specific or a cross-domain standard.		lity	vocabularies are used.	
[Requirement 67]	Controlled values used for the description of resources must be identified by identifiers that are unique, machine readable and persistent.	Property	Identifiers	Unique identifiers (e. g. GND-URI) are provided only for those terms applied in the cross-collection databases. The vocabularies, used within the local developed databases and in the proprietary systems, can only be identified by the term itself, the identifiers are ignored.	(✓)
[Requirement 68]	The used controlled vocabularies must be available in machine readable form.	Value	Metadata	Mainly, this is the case concerning the vocabularies used within the cross-collection databases, but concerning the other vocabularies this requirement is fulfilled only conditionally.	(✓)
[Requirement 69]	The used controlled vocabularies must be available open access.	Value	Rights	Mainly, this is the case concerning the vocabularies used within the cross-collection databases, but concerning the other vocabularies this requirement is fulfilled only conditionally.	(✓)
[Requirement 70]	Synonym appellations must be merged using authorities.	Value	Metadata	This is possible only in those cases, where a singular terms can be linked by using an unique identifier.	(✓)